

Indexación y códigos en bases de datos bibliográficas

JEAN-MARC BUIGUÈS
Université Michel de Montaigne Bordeaux3
JEAN-MARC.BUIGUES@U-BORDEAUX3.FR

INTRODUCCIÓN

En la primera parte de esta exposición presentaré a grandes rasgos el proyecto de bases de datos denominado NICANTO (Buiguès, 2011; Dedieu, 1997: 325-336). A continuación, me centraré en los problemas de indexación y codificación de datos en bases de investigación en las humanidades digitales, apoyándome en ejemplos concretos y en la experiencia acumulada en NICANTO. Sucesivamente evocaré la indexación y codificación de fechas, de lugares y de materias y la metodología seguida en el análisis de la sociología de los autores.

1. PRESENTACIÓN DE NICANTO

El proyecto NICANTO nació hace diecinueve años en la Maison des Pays Ibériques de la Universidad Michel de Montaigne Bordeaux3, un centro de estudios que funcionaba como una UMR (Unidad mixta de Investigación cuyo nombre era TEMIBER) que asociaba investigadores del CNRS y de la Universidad. Iniciado por tres investigadores: Jean-Pierre Dedieu (Director de investigaciones el CNRS) el añorado profesor François Lopez (Catedrático de la Universidad Michel de Montaigne Bordeaux3) y su servidor, en aquel momento recién nombrado Profesor Titular en dicha Universidad. Unimos nuestra experiencia en la investigación. François Lopez era el que mejor conocía la historia de la edición y de la imprenta españolas, también proyectaba informatizar la *Bibliografía de autores españoles del siglo XVIII* de Francisco Aguilar Piñal (Aguilar Piñal, 1981-2002) y los anuncios de impresos publicados en el siglo XVIII en la *Gaceta de Madrid*; Jean-Pierre Dedieu tenía una gran experiencia en bases de datos, experiencia acumulada en particular en la elaboración de la base FICHOZ sobre la alta administración española en la época moderna; yo llegaba con la experiencia adquirida durante la redacción de mi tesis doctoral (Buiguès, 1991) asentada en una base de datos sobre inventarios *post mortem*, testamentos, etc. y en la cual dedicaba una parte al análisis de los lectores y de las lecturas de los leoneses en el siglo XVIII a partir del estudio de sus bibliotecas –para elaborar un proyecto común bautizado “NICANTO” en homenaje a Nicolás Antonio.

JEAN-MARC BUIGUÈS

«Indexación y códigos en bases de datos bibliográficas»

Humanidades Digitales: desafíos, logros y perspectivas de futuro. Janus, Anexo 1 (2014), pp. 123-135

El proyecto NICANTO se ideó como un sistema de bases de datos relacionadas entre ellas (ver Fig. 1). El núcleo central lo constituyó la primera base "Gnledici" en la que se informatizaron los ocho primeros tomos (los dedicados a obras con autoría) de la *Bibliografía de autores españoles del siglo XVIII*. Una segunda base, "Gnlrespo" que reunía a los autores fue creada a partir de Gnledici. La informatización de inventarios de bibliotecas privadas e institucionales españolas del siglo XVIII se empezó en dos otras bases. Una quinta base sirvió para identificar las fuentes utilizadas en las bases sobre inventarios de bibliotecas y reunir informaciones cualitativas sobre los poseedores de libros (calidad, oficios y cargos, niveles de fortuna, presencia de cuadros, etc.). La penúltima base fue la dedicada a la informatización de los anuncios de impresos publicados en el siglo XVIII en la *Gaceta de Madrid*, siendo la última base denominada "Censura" la que pretendía informatizar los índices inquisitoriales del siglo XVIII.

Actualmente la base Gnledici totaliza unas 26.000 fichas de 80 campos cada una, la base Gaceta, unas 19.000 fichas con 62 campos, las bases de inventarios de bibliotecas, unas 15.000 fichas con 60 campos, Gnlrespo, unas 7.600 fichas con 24 campos.

En realidad, el proyecto NICANTO se puso en marcha para poder desarrollar varios ejes de investigación:

- Historia de la edición y de la imprenta¹.
- Historia de los autores².
- Publicidad del impreso (anuncios de la *Gaceta de Madrid*)³.
- Lectores y lecturas, bibliotecas privadas y bibliotecas institucionales⁴.
- Evolución de la censura en el siglo XVIII.

Queda patente que a lo largo de estos casi 20 años de investigación, el proyecto inicial sufrió evoluciones debidas entre varios factores principalmente a la evolución de la estructura institucional de la investigación (Desaparición de la UMR TEMIBER), a las vivencias personales de los investigadores que constituyeron el núcleo fundador del proyecto, a las evoluciones técnicas en el mundo de la informática y de Internet, etc. Entre los ejemplos más significativos de esta evolución están la participación de NICANTO desde hace cuatro al proyecto PHEBO (Poesía Hispánica en el Bajo Barroco⁵) y más recientemente su participación a Dialogyca (Biblioteca Digital de Diálogo Hispánico⁶).

Sin embargo, las investigaciones llevadas a cabo permitieron enfrentarse con toda una serie de problemas que hoy en día siguen siendo fundamentales para quien desea basar su investigación en bases de datos relacionales como por ejemplo la homogeneización de los datos o los métodos que permitan acelerar

¹ Unos primeros resultados se publicaron en Infantes, López y Botrel (2003) y más precisamente en Buiguès (2003: 303-316) y Buiguès (2003: 317-327).

² El lector encontrará un ejemplo de la investigación sobre esta temática en Buiguès (2003: 292-302)

³ Un ejemplo de la investigación sobre este eje en Jean-Marc Buiguès (2013: 33-47).

⁴ Ejemplos de la investigación que corresponde a este eje en Infantes, López y Botrel (2003: 424-433) y Buiguès (2008: 361-367).

⁵ Ver PHEBO, <<http://phebo.es/es/presentacion>>.

⁶ Ver Dialogyca BDDH (Biblioteca Digital de Diálogo Hispánico), <http://pendientedemigracion.ucm.es/info/dialogycabddh/f_presentacion.html>.

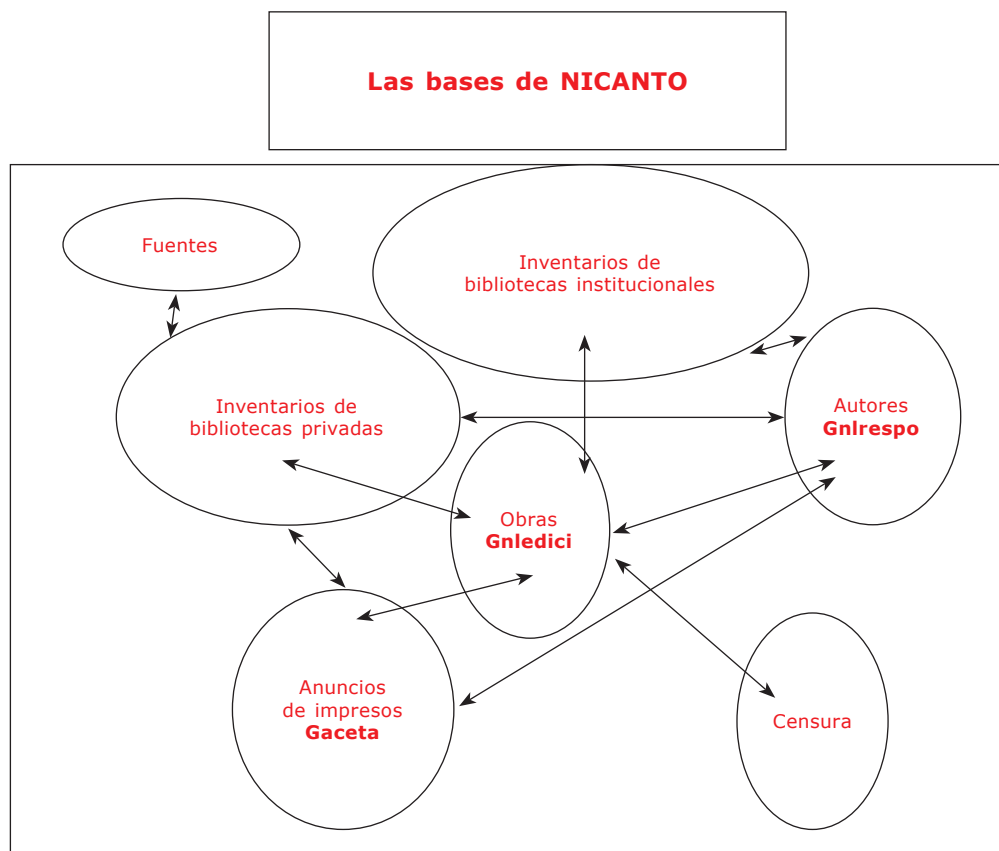


Fig. 1.

las búsquedas. Entre los procedimientos que se utilizaron para resolver dichos problemas presentaré a dos de ellos: la indexación y la codificación.

2. INDEXACIÓN Y CODIFICACIÓN DE DATOS EN NICANTO

Antes de presentar algunos ejemplos de indexación y de codificación⁷ quizás sea necesario dar unas cuantas definiciones. Indexar es en informática, la "acción de ordenar registros empleando índices. Un ejemplo son las bases de datos, que emplean uno o más índices para identificar registros", siendo el índice en una base de datos "una o más claves que identifican un único registro (fila o renglón). Los índices hacen más fácil encontrar y ordenar los registros"⁸. El término codificación es el "Proceso de organización de la información en

⁷ Es de señalar que "indexación" y "codificación" no figuran todavía con estos significados en el DRAE.

⁸ Ambas definiciones provienen del *Diccionario de informática*, <<http://www.alegsa.com.ar/Dic/indexar.php>>.

categorías, mediante el cual se asignan códigos para introducir, almacenar y recuperar datos”⁹.

Estos procesos no son sino procesos de transformación de la información de la fuente primaria, o dicho de otro modo procesos que participan en distintos momentos de la investigación que son la entrada, el almacenamiento y la búsqueda de la información.

2.1. Ejemplos de indexación en NICANTO

Un primer ejemplo de indexación es el que ofrece el campo “Nº_Edición” en la base Gnedici. Este campo es alfanumérico. Empieza por letras seguidas por dígitos. Veamos unos cuantos ejemplos: “BEGUE1”, “AP-A01396”¹⁰. En el caso de “AP-A01396”, AP indica que la fuente de la ficha es la *Bibliografía de autores españoles del siglo XVIII* de Francisco Aguilar Piñal; la letra “-A” (“AP-A01396”), el tomo 1 y “01396” el número de identificación de la ficha en la *Bibliografía*. Para “BEGUE1”, “BEGUE” indica que la fuente es la “Relación de la poesía española publicada entre 1648 y 1750” (Bègue, 2010) y “1” que es la primera obra mencionada en dicha relación. Se hubiera podido utilizar para este campo “Nº_Edición” únicamente un número pero el sistema de indexación elegido permite buscar en función de la fuente. Por ejemplo una búsqueda en dicho campo con “AP-B” me da todas las fichas del segundo tomo de la *Bibliografía* lo que una indexación meramente numérica no hubiera permitido.

Del mismo modo en la base *Gaceta*, el campo “Número” se rellena de la manera siguiente “GM-004093” en que “GM” significa la *Gaceta de Madrid* y “004093” el número de la ficha por orden de entrada de datos. La mención de “GM” permite prever la futura y/o posible informatización de anuncios de impresos de otros periódicos. Por ejemplo si se tratase de anuncios del *Memorial literario* se podría rellenar el campo con “ML”.

2.2. Codificación de fechas

Para establecer la cronología de la edición en la España del siglo XVIII fue necesario profundizar la utilización de las fechas tal como se mencionan en las fuentes primarias. Las fechas presentes en las fuentes primarias ofrecen según las fuentes una gran variedad desde su total ausencia como en el caso de los romances populares publicados en el siglo XVIII hasta la pormenorizada mención del año, mes y día como lo ofrecen los anuncios de impresos de la *Gaceta*. Si examinamos por ejemplo las fechas de publicación en Gnedici uno de los problemas que se plantean es por ejemplo el de las obras con varios volúmenes publicadas durante varios años. La dieciséis volúmenes de la “Synopsis histórica cronológica de España. parte primera que comprende sus sucesos desde la creación de el mundo hasta el nacimiento de Jesu Christo ...: formada de los autores seguros y de buena fee / por don Juan de Ferreras ...impresa en Madrid por Francisco de Villa-Diego: a costa de Diego Lucas

⁹ Según definición de *Los diccionarios y las enciclopedias sobre el Académico*, <http://www.esacademic.com/dic.nsf/es_mediclopedia/30490/>.

¹⁰ Al proyecto inicial que sólo preveía la informatización de la *Bibliografía* de Aguilar Piñal se han añadido otros datos procedentes de varias fuentes: bibliografías impresas, recursos digitales, etcétera.

Codificación cronológica

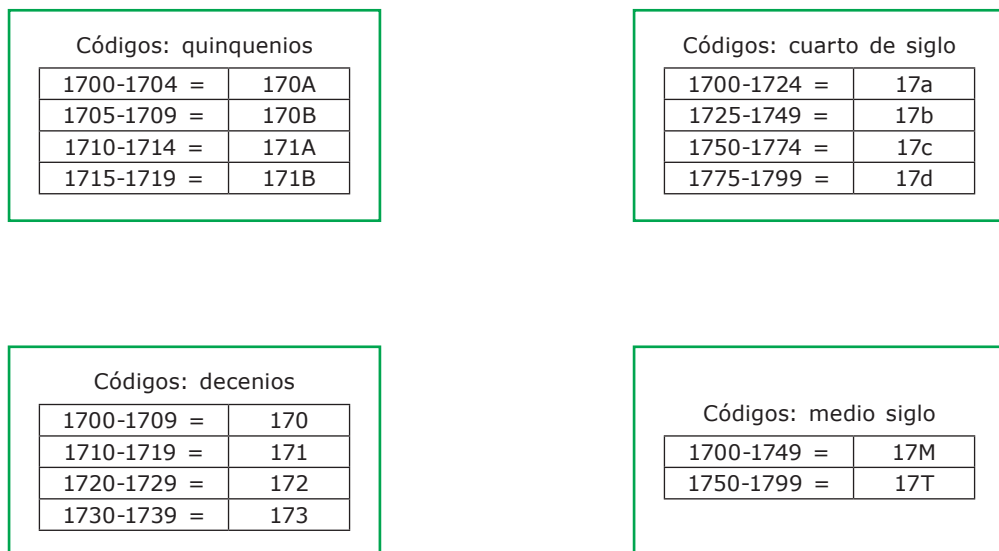


Fig. 2.

Ximenez¹¹ se publicaron entre 1702 y 1727. El campo "Fecha_o" se rellenó así: "1702-1724". Cuando el investigador hace una búsqueda por fechas en el campo "Fecha_o" sólo saldrá la ficha si tecllea "1702" o "1727". Es menester entonces crear campos de codificación de las fechas. La figura 2 indica cuatro modalidades de codificación de las fechas: por quinquenios, por decenios, por cuartos de siglo o por mitades de siglo¹². En el ejemplo de la "Synopsis" de Juan de Ferreras se rellenará el campo "medio siglo" con "17M" (ver Fig. 2). En función de la precisión de la fecha de edición –cuando la obra carece de fecha de publicación las fechas que servirán para rellenar el campo ""Fecha_o" pueden ser por ejemplo las que corresponden al período de actividad del impresor, y por consiguiente los campos correspondientes de codificación de las fechas, etc.– se rellenarán todos los campos o solamente algunos. La codificación de las fechas permite por ejemplo trazar la curva de la producción de obras nuevas en castellano durante el siglo XVIII como lo evidencia el gráfico de (Fig. 3) (Infantes, López y Botrel, 2003). En este caso por la abundancia de datos fue posible utilizar los

¹¹ Datos que provienen de la *Bibliografía* (tomo 3, nº 3781). Se respeta la ortografía de la fuente.

¹² Se crearán cuatro campos: "Fecha_quinquenio", "Fecha_decenio", "Fecha-_cuarto" y "Fecha_mitad".

**Producción de obras nuevas en castellano
(curva anual y quinquenal)**

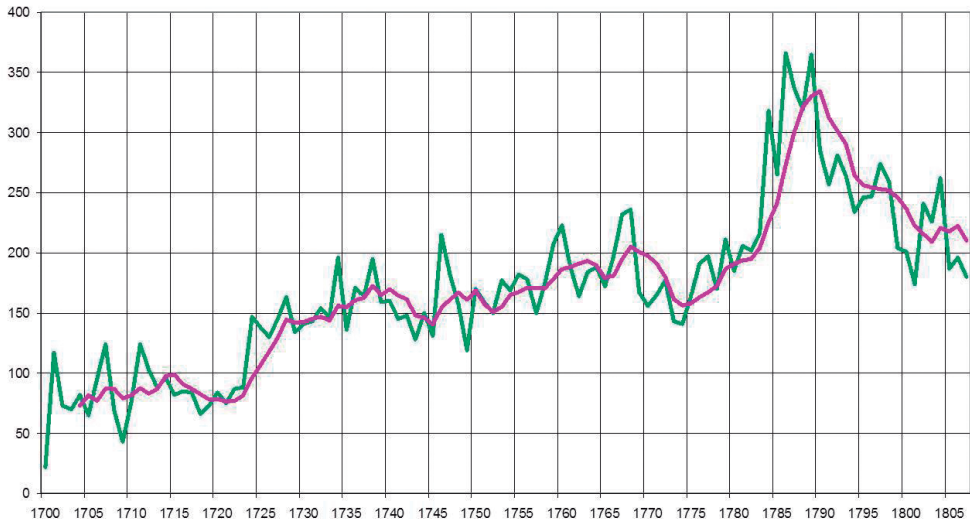


Fig. 3.

campos "Fecha_o" y "Fecha_quinquenio" porque había datos suficientes como para establecer las dos curvas. En el caso por ejemplo del romancero popular del siglo XVIII, la ausencia muy importante de la mención del año de edición sólo permite y ni siquiera siempre rellenar el campo "Fecha-mitad" para medir las evoluciones del género sólo entre la primera y la segunda mitad del siglo.

2.3. Codificación de lugares

En la base Gnedici se creó un campo "Lugar_O" para el lugar de edición de la obra tal como aparece en la fuente. Sin embargo no sólo las variantes ortográficas (B/V, etc.) sino también la presencia de los nombres de las ciudades en latín y en castellano (pero también a veces en otros idiomas), dificulta las búsquedas. Por ejemplo para la ciudad de Lyon, el investigador tiene que buscar sucesivamente por "Lyon", "León" y Lugduni". Además en el caso de la búsqueda por "León" aparecerán no sólo fichas de "León de Francia" según la terminología de la época sino también las de la castellana ciudad de León y las de la Isla de León, ciudad ésta que por otra parte aparece también en otras fichas designada por San Fernando. Para homogeneizar estos datos, evitar confusiones y facilitar las búsquedas, se creó un campo "Lugar_H" donde se pone el nombre moderno: "Lyon", "León" para la ciudad castellana y "San Fernando" para la Isla de León. Así el investigador que busque por "Lyon" podrá localizar en una sola búsqueda todas las fichas de Lyon, León y Lugduni.

Por otra parte, unos de los ejes de la investigación era establecer la geografía de la edición lo que necesitó la creación de un campo de codificación de esta información. Por ejemplo para medir la producción del impreso poético español publicado en Europa entre 1650 y 1699, hacía falta un campo de codificación

que permita distinguir la producción francesa de la italiana por ejemplo. Se creó para dicho fin el campo "Cod_Lugar" con tres dígitos. Por ejemplo en el caso de "EFA" la letra "E" corresponde a Europa, la letra "F" a Francia y la "A" designa París. Se puede con este campo saber pues cuántas obras de poesía española se publicaron en Europa, en Francia o en París. Esta codificación de los lugares permitió establecer el mapa de las ediciones de impresos poéticos españoles en Europa (Buiguès, 2013) y (López, Sáez y Sánchez, en prensa) (ver Fig. 4).

El campo "Cod_Lugar" permite también llevar a cabo estudios regionales como por ejemplo medir la producción gallega en el siglo XVIII. En el cuadro siguiente la letra "I" corresponde a Galicia, mientras las letras "A", "B", "C" y "D" a las ciudades de respectivamente La Coruña, Santiago de Compostela, El Ferrol y Orense.

	<i>Ediciones</i>	<i>%</i>	<i>Cod_Lugar</i>
Orense	2	1,0	PID
El Ferrol	2	1,0	PIC
La Coruña	23	11,2	PIA
Santiago de Compostela	179	86,9	PIB
Total Galicia	206		

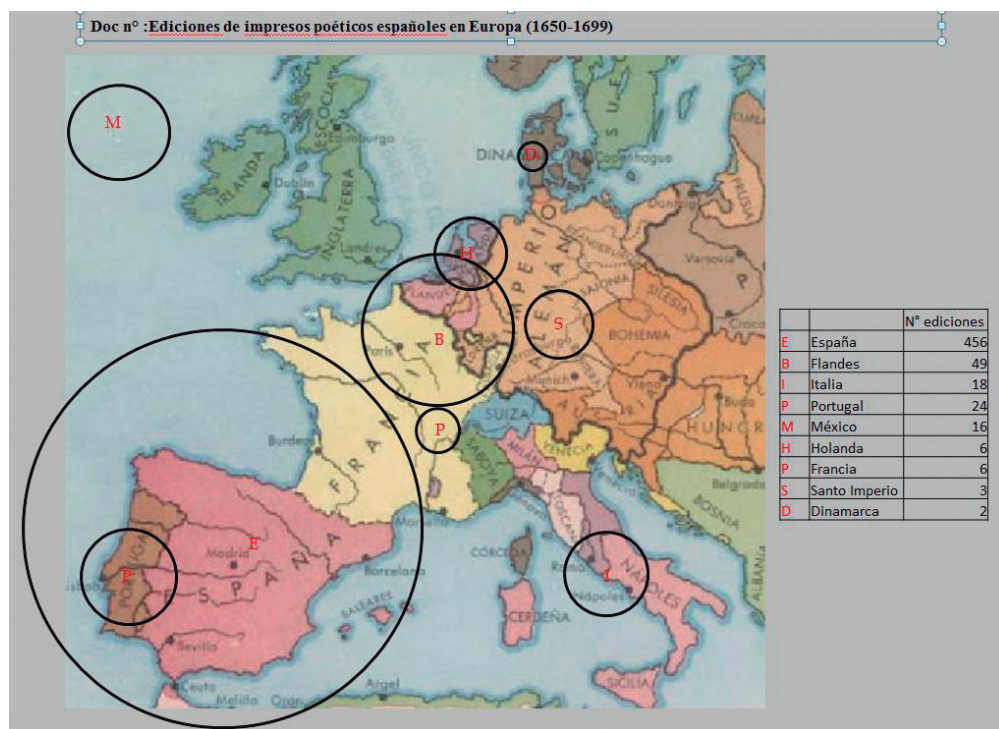


Fig. 4.

Otra base importante de NICANTO es la base que reúne los anuncios de impresos de la *Gaceta de Madrid* entre 1697 y 1807. Entre los datos que ofrecen los anuncios figura el lugar de venta del impreso que puede ser un local de profesionales (libreros, impresores) pero también la casa del autor, la portería de un convento, una tienda de comestibles, etc. Además de la variedad tipológica de los puestos de venta (librero, portería de convento, etc.) un mismo puesto puede mencionarse bajo distintas formas. Por ejemplo, los propios anuncios de la *Gaceta* se anuncian a veces "Donde la Gazeta" y otras "Donde estas relaciones se hallará". Otro elemento importante es que los anuncios no son sólo de Madrid, los hay también de otras ciudades. Fue necesario crear un campo de codificación y homogeneización de estos datos: el campo "CodPuestoVenta". A continuación se dan algunos ejemplos de esta codificación:

Md180 = Puesto de Francisco Robledo, Puerta de Guadalajara
 Md181 = En la imprenta de Joseph Rodriguez, Calle del Carmen
 Vd3 = Casa de la viuda de Santánder
 VI8 = Casa de Manuel Cabero, calle de Campaneros

Las letras "Md", "Vd" y "VI" corresponden a Madrid, Valladolid y Valencia, los números a un lugar de venta. Este campo permite medir la importancia de cada ciudad en cuanto a la publicidad del impreso, pero también de ver para cada ciudad quiénes son los que anuncian en la *Gaceta*. También se creó un campo de codificación del tipo de lugar de puesto de venta llamado "CodTipoPuesto" en que por ejemplo "P" corresponde a un anuncio de portería de convento, "I" al de una imprenta, "L" al de un librero, "T" al de una tienda, etc. Utilizando este campo se puede por ejemplo para Madrid, evaluar el peso de los anuncios de profesionales del libro frente al de los no profesionales.

2.4. Sociología de autores

La base sobre autores, Gnlrespo, proporcionaba la posibilidad de hacer un estudio de la sociología de los autores dieciochescos. Para este fin se crearon varios campos de codificación de los cuales sólo mencionaré tres. "CATEG1" permite distinguir el estamento (nobleza, clero o estado llano), pero también por ejemplo para los eclesiásticos si pertenecen al clero regular o secular, de qué orden son o qué función o cargo desempeñan. Por ejemplo si en el campo CATEG1 tecleo "R-SJ" aparecen todas las obras cuyos autores son jesuitas o sea 373 autores que publicaron 7593 obras. Del mismo modo el campo CATEG2 permite hacer búsquedas a partir de los oficios de los autores: así "F-AA" permitirá localizar a los miembros de los Consejos Reales, "F-MI" a los ingenieros de marina y "M-AP" a los pintores. Sin embargo una búsqueda con sólo "M" hubiese permitido seleccionar a los autores que ejercitaban oficios mecánicos y artísticos, "M-A" a únicamente los artistas. El campo CATEG3 es el que permite hacer búsquedas para los miembros de Academias con el mismo sistema de informatización jerarquizada en función del rango de cada letra: por ejemplo si busco por "A-HV", la "A" corresponde a un miembro de una academia, la "H" indica que se trata de la Real Academia de la Historia y la "V" que el autor desempeña el cargo de director.

2.5. Codificación de Materias

Cuando se creó la base Gnlledi uno de los ejes de investigación previstos era el estudio de la producción impresa en función de las materias. Una solución

<p>Listado de materias</p>	<p>Academias, Antologías, Artículos periodísticos, Atribuciones, Autobiografías, Bibliografías, Biografías, Cancioneros, Cartas, Compilaciones, Crónicas, Documentos legales, Estudios, Facsimiles, Genealogías, Glosas, Historias de escritoras, Justas, Manuscritos, Menciones coetáneas, México, Obras dramáticas, Obras generales, Obras perdidas, Páginas web, Partidas de bautismo, Pliegos sueltos, Poemas, Portugal, Preliminares, Profesiones religiosas, Prosa de Ficción, Prosa de pensamiento, Prosa religiosa, Relaciones, Siglo XIV, Siglo XV, Siglo XVI, Siglo XVII, Siglo XVIII, Tesis doctorales, Testamentos, Traducciones, Tratados</p>	7. TABLA DE TÉRMINOS ADMITIDOS DE "MATERIAS Y TEMAS SECUNDARIOS"		
<p>Otras materias de búsqueda</p>	<p>Nombres de autoras Siglo en que escribieron (números romanos) Género de las obras Estudios o ediciones</p>	<p>Adivinación Agricultura Alimentación Alquimia América Amor Anatomía Antropología Arbitrismo Aritmética Armas y letras Arqueología Arquitectura Arte Artes de bien morir Artes de gobernar Artes de la guerra Artes de navegar Artilería Ascética Astrología Astronomía Biología Botánica Catequesis Caza Cirugía Ciudades Comercio Controversia religiosa Cosmografía Cosmología Cristianismo Crítica de costumbres Crítica social Danza Derecho Dialéctica Dietética Economía Edades de la vida Educación</p>	<p>Emblematología Enseñanza de lenguas Escatología Escritura Espiritualidad Estados civiles Estamentos Estética Ética Enqueta Emografía Farmacia Filografía Filosofía Filosofía moral Filosofía natural Finanzas Física Geografía Geología Geometría Gramática Guerra Heraldica Herejía Historia Honor Iconografía Ingeniería Islam Judaísmo Juegos Lengua Linajes Literatura Lógica Locura Magia Máquinas Matemáticas Mecánica</p>	<p>Medicina Melancolía Metafísica Metalurgia Mística Mitología Mujer Música Naturaleza Navegación Negocios Numismática Odontología Oficios Óptica Ordenes militares Ordenes religiosos Ortografía Países Profesiones Pescas Pintura Poesía Política Psicología Química Religión Retórica Sucesos particulares Superstición Teatro Teología Teoría de la historia Teoría de los saberes Territorios Textos sagrados Traducción Varia lección Veterinaria Vida de corte y de aldea</p>
<p>Siglo de una autora</p>	<p>De forma general se consigna el siglo en el que vive y escribe, no el de publicación de su obra. Autoras que viven entre dos siglos, se incluyen los dos</p>	<p>DIALOGYCA BDDH, BIBLIOTECA DIGITAL DE DIÁLOGO HISPÁNICO, Manual de procedimiento de trabajo normalizado, (3ª versión febrero 2009), p.125</p>		
<p>Búsqueda combinada de materias</p>	<p>Se pueden combinar varios conceptos Se pueden usar ventanas diferentes para cada concepto o ponerlos en la misma ventana sin separación entre ellos</p>			
<p>Menciones coetáneas/BIESES</p>	<p>Se refiere a citas y menciones de las autoras, hechas en su propio tiempo o poco después</p>			
<p>BIESES, BIBLIOGRAFÍA DE ESCRITORAS ESPAÑOLAS, pantalla de ayuda para realizar búsquedas en el campo materias, http://62.204.193.244:8080/bieses/ayuda/bieses.htm#busquedapormaterias</p>				

Fig. 5.

posible hubiera sido crear un campo "Materia" y establecer un listado de las palabras que pudieses servir para rellenar dicho campo. Es por ejemplo el sistema que suelen utilizar las bases de datos bibliográficas como la de Biblioteca Nacional de España o la del Catálogo Colectivo del Patrimonio Bibliográfico Español. En regla general las búsquedas por materias en este tipo de bases no suelen dar resultados muy completos por distintas razones: por ejemplo en el Catálogo Colectivo, este campo se rellenó según criterios diferentes en función de la institución que conserva el ejemplar de una obra (más de setecientas instituciones) pero también con criterios que cambiaron a lo largo del tiempo para una misma institución. En el caso de bases especializadas elaboradas por grupos de investigación cada base elabora en función de sus ejes de trabajo soluciones distintas como se puede ver en las dos magníficas bases Dialogyca y Bieses (ver Fig. 5). La solución en Gnledici fue distinta: se crearon varios campos para las materias y formas. En un campo llamado MAT1 se establecieron, a partir de una clasificación de las materias según la clasificación elaborada para las bibliotecas del Antiguo Régimen por el historiador François Furet (Furet, 1965: 3-32)¹³, cinco grandes categorías. "A" = Religión, "B" = Derecho, "C" = Historia, "D" = Ciencias y Artes, "E" = Bellas Letras. Cada categoría se subdividía. Por ejemplo la categoría "Ciencias y Artes" se componía, siempre basándose en la clasificación

<p><u>D2.—ARTES</u></p> <p><u>B: Artes mecánicas</u></p> <p>1: Relojes</p> <p>2: Madera</p> <p>3: Seda</p> <p>4: Platería</p> <p>5: Hierro</p> <p>6: Cobre</p> <p>7: Lana</p> <p>8: Pirotecnia</p> <p>9: Fuego</p> <p>10: Vidrio</p> <p>11: Papel</p> <p>12: Cuero</p>	<p><u>D2.—ARTES</u></p> <p><u>A: Artes liberales</u></p> <p>1: Música</p> <p>2: Pintura</p> <p>3: Escultura</p> <p>4: Dibujo</p> <p>5: Grabado</p> <p>6: Arquitectura</p> <p>7: Arte militar</p> <p>8: Arte de escribir</p> <p>9: Decoración</p> <p>11: Danza</p> <p>12: Ópera</p>
--	--

Fig. 6.

de François Furet, de los bloques siguientes: D1.—Ciencias, D2.—Artes, D3.—Agricultura y agronomía, D4.—Juegos de destreza y pasatiempos, D5.—Economía y política, D6.—Diversos. Del mismo modo cada una de estas subcategorías se dividía en otros bloques. Por ejemplo "D2" que corresponde a Artes se subdividía a su vez en "D1B" para las artes mecánicas y en "D2A" para las artes liberales. Por fin estas dos subcategorías daban lugar a una última caracterización con el añadido de cifras: por ejemplo, "D2A8" identificaba las obras que versaban sobre el arte de escribir (ver Fig. 6). Este sistema de codificación permitía medir por ejemplo el peso de la producción de obras científicas en el conjunto de la producción pero también según lo que el investigador buscaba localizar por ejemplo a todas las obras que trataban de cirugía.

Sin embargo, como se mencionaba en la primera parte de este trabajo, las líneas de investigación fueron cambiando con el transcurso de los años lo que hizo que nuevos campos para las materias tuvieron que ser creados. Dentro de las últimas líneas de trabajo está la dedicada a la poesía posbarroca dentro del equipo que dirige Pedro Ruiz Pérez¹⁴. Para estudiar este tipo de impreso se han creado campos específicos. Por ejemplo un campo "Protagonista" en que se pone la identidad del sujeto sobre que versa el poema: "Santo Tomás de Aquino", "Carlos III", "doña Rosa de Guzmán Silva, dignísima hermana del duque de Medina Sidonia", etc.; un campo "Acontecimiento" en que se especifica por ejemplo "Nobleza, bodas", "Academia poética, certamen", "polémica teatral" y otros que no se mencionan aquí. En esta primera fase del trabajo se reelaboran tanto informaciones que da la fuente primaria como informaciones añadidas por el investigador. Sin embargo para tener una visión de conjunto será indispensable crear campos de codificación para por ejemplo los acontecimientos. Si se quiere analizar los poemas que tratan de episodios de la vida de la familia real y los de la nobleza es posible crear un campo "Cod_Acontecimiento" con un sistema jerarquizado de la información que podría ser el siguiente: la primera

¹⁴ "Poesía del período posbarroco (1650-1750): repertorios y categorías" (FFI2008-03415-E/FILO del MICINN).

letra indicaría si el protagonista pertenece a la familia real, etc. la segunda el tipo de acontecimiento, boda, nacimiento, defunción, etc. Así "RB" indicaría que se trata de las bodas de un miembro de la familia real, "NN, del nacimiento de un noble, etc. De este modo el investigador ahorraría tiempo y encontraría las debidas respuestas a las preguntas formuladas en su investigación.

CONCLUSIÓN

El primer propósito de este trabajo fue exponer el camino seguido en Nicanto para ir de la constitución de una base que hubiera podido ser meramente bibliográfica hasta la elaboración de una base de investigación. El elenco de ejemplos de indexación y de codificación presentados aquí demuestra que sin ellos hubiera sido casi imposible llegar a los resultados publicados. Por ejemplo por mucho que uno manejara las 23.000 fichas de la *Bibliografía* de F. Aguilar Piñal, le hubiera sido casi imposible establecer una curva de la producción del impreso español en el siglo XVIII o saber cuántas obras fueron publicadas por los jesuitas. La informatización de la monumental cantidad de datos recogidos por F. Aguilar Piñal, su indexación y codificación sí que lo permitieron.

Exponer el camino recorrido, señalar unas cuantas técnicas de homogeneización y codificación de datos apuntando cómo en definitiva fueron los distintos ejes de investigación y los propios investigadores los que definieron estos procesos era lo que se pretendía aquí.

Mas no sólo: me pareció también importante reflexionar o solicitar una reflexión sobre la distinción entre lo que son las bases de datos bibliográficos de consulta pública que no fueron, ni son todavía ni tampoco tienen por qué serlo, elaboradas para la investigación. Estas bases cuya grandísima utilidad no hace falta subrayar proporcionan una información primaria que viene a completar la información más tradicional que da el impreso o el manuscrito: son unas de las fuentes que permiten la creación de un segundo tipo de bases que elabora la investigación. En el caso de las bases de investigación como Nicanto, los investigadores van a añadir toda una serie de informaciones a las que proporcionan las fuentes primarias. Entre estas informaciones añadidas figuran como acabamos de verlo no sólo las que provienen de la necesidad de homogeneizar datos multiformes sino también los procesos de codificación de la información que permiten, sobre todo si la base cuenta con un gran número de fichas, contestar de manera rápida y eficaz a las problemáticas definidas por cada eje de investigación.

El vaivén constante entre la información que proporcionan las fuentes primarias y las que establece la investigación con la también necesaria posterior codificación son en definitiva lo que distingue una base de consulta abierta a todos y una base de investigación. Lo que no significa que una base elaborada para la investigación no pueda también pasar a ser base de consulta pública. El problema es la relación entre ambas o dicho de otro modo cuál es la información de la base de investigación que se da a ver libremente al público. La respuesta sea quizás compleja. Si algunas bases de investigación pueden dejar libre acceso a la totalidad de sus fichas es sin duda porque no permiten búsquedas elaboradas en función de cada nuevo investigador que consulte la base dado que suelen llevar pocos campos y ninguno codificado, y sobre todo porque

en los distintos campos la información no se ofrece de manera jerarquizada. Ciertos tipos de búsquedas como las evocadas anteriormente seguirían sin embargo siendo siempre posibles en estas bases pero no serían nada fáciles y llevarían mucho tiempo. Por otra parte poner a libre disposición todo cuanto fue elaborado por un investigador para su investigación tampoco es deseable porque permitiría a otros investigadores utilizar estas codificaciones para sus propias investigaciones sin que sea siempre posible comprobar la fuente. Esta información añadida en las bases de datos de investigación a través de los procesos de codificación es finalmente una nueva modalidad de la creación, y como tal forma parte del derecho a la propiedad intelectual. La solución consistiría sin duda a definir perfiles de usuarios que diesen acceso a distintos niveles de información.

BIBLIOGRAFÍA

- Aguilar Piñal, Francisco, *Bibliografía de autores españoles del siglo XVIII*, Madrid, CSIC, 1981-2002, 10 tomos.
- Bègue, Alain, "Relación de la poesía española publicada entre 1648 y 1750", en Aurora Egido y José Enrique Laplana (eds.), *La luz de la razón. Literatura y Cultura del siglo XVIII. A la memoria de Ernest Lluch*, Zaragoza, Institución "Fernando el Católico" (Col. Actas), 2010, pp. 399-477.
- Buiguès, Jean-Marc, *La ville de Léon et sa juridiction. Economie, société et mentalités en Vieille Castille au XVIII^{ème} siècle*, Tesis dirigida por Jacques Soubeyroux, Universidad de Montpellier, 1991.
- Buiguès, Jean-Marc, "Evolución global de la producción", en Víctor Infantes, François López y Jean-François Botrel (coords.), *Historia de la edición y de la lectura en España 1472-1914*, Madrid, Fundación Germán Sánchez Ruipérez, 2003, pp. 303-316.
- Buiguès, Jean-Marc, "Las materias: tradición y modernización", en Víctor Infantes, François López y Jean-François Botrel (coords.), *Historia de la edición y de la lectura en España 1472-1914*, Madrid, Fundación Germán Sánchez Ruipérez, 2003, pp. 317-327.
- Buiguès, Jean-Marc, "Géographie de l'édition des livres de l'abbé de la collégiale de Villafraanca del Bierzo (1795)", en Philippe Meunier y Edgard Samper (coords.), *Mélanges en hommage à Jacques Soubeyroux*, Saint-Etienne, Celec, 2008, pp. 361-367.
- Buiguès, Jean-Marc, "NICANTO. Obras impresas de escritores del siglo XVIII", en María Casas del Álamo, Germán Redondo Pérez y Sara Sánchez Bellido (eds.), *Actas del Seminario Internacional sobre Bibliotecas Digitales y Bases de datos Especializadas para la investigación en Literaturas Hispánicas* (BIDESLITE), Madrid, IUMP (UCM), 2013, <http://eprints.ucm.es/21207/1/Actas_Seminario_BIDESLITE.pdf>.
- Buiguès, Jean-Marc, "Les annonces d'imprimés littéraires dans la presse espagnole: 1697-1737", en Florence Boulerie (ed.), *La médiatisation du littéraire dans l'Europe des XVII^e et XVIII^e siècles*, Biblio 17, vol. 205, suplemento a Papers on French Seventeenth Century Litteratue, Tübingen, Narr/Verlag, 2013, pp. 33-47.
- Buiguès, Jean-Marc, "El impreso poético posbarroco: centro y periferia, impresores y formas materiales", *Versants*, nº 60:3 (fascículo español) (2013).
- Dedieu, Jean-Pierre, "El sistema Nicanto", *Bulletin Hispanique*, 99-1 (1997), pp. 325-336.
- Furet, François, "La 'librairie' du royaume de France au XVIII^e siècle", *Livre et société dans la France du XVIII^e siècle*, Paris, Mouton, 1965, pp. 3-32.
- Infantes, Víctor; Lopez, François y Botrel, Jean-François (coords.), *Historia de la edición y de la lectura en España 1472-1914*, Madrid, Fundación Germán Sánchez Ruipérez, 2003.



RESUMEN

El presente artículo presenta en su primera parte las bases de datos relacionales que constituyen el proyecto NICANTO del equipo de investigación AMERIBER de la Universidad de Burdeos III así como su evolución desde su creación hace veinte años. El proyecto abarca un amplio abanico de líneas de investigación: historia de la edición y de la imprenta, sociología de los autores, publicidad del impreso en la prensa, lectores y lecturas, bibliotecas privadas y bibliotecas institucionales, evolución de la censura en el siglo XVIII, poesía posbarroca (1650-1750), diálogos del siglo XVIII. El marco cronológico lo constituyen los siglos XVIII y XIX, el marco espacial España y Cuba. Las distintas bases contabilizan unas 75.000 fichas.

La segunda parte del artículo expone una serie de métodos elaborados en función de las distintas líneas de investigación. Estos métodos se centran en las nociones de indexación y de codificación de los datos informatizados a partir de fuentes primarias. Los ejemplos desarrollados corresponden tanto a indexación de la localización de las fuentes primarias como a codificaciones de campos de fechas, de lugares, de autores o de materias.

Por fin se sugiere una reflexión en torno a la distinción entre base de datos de consulta pública y base de datos para investigación.

Palabras claves: Bases de datos relaciones, España, siglo XVIII, edición, bibliotecas, codificación de datos.

ABSTRACT

This article presents in its first part the relational databases that constitute the project NICANTO of the AMERIBER research of the University of Bordeaux III team as well as its evolution since its creation twenty years ago. The project covers a wide range of research interests: history of editing and printing, sociology of the authors, advertising of the printed in the press, readers and readings, private libraries and institutional libraries, evolution of censorship in the 18th century, poetry posbarroca (1650-1750), dialogues of the 18th century. The chronological framework is the 18th and 19th centuries, the spatial context Spain and Cuba. The different foundations accounted for a 75,000 tokens.

The second part of the article exposes a number of methods elaborated according to the different lines of research. These methods focus on notions of indexing and encoding of the data computer from primary sources. Developed examples correspond to index the location of primary sources and encodings of fields of dates, places, authors or subjects.

Finally it is suggested that a reflection on the distinction between inspection database and database for research.

Keywords: Database relations, Spain, 18th century, Edition, libraries, data encryption.

